

# Los tesauros y las ontologías en la Biblioteconomía y la Documentación

## Silvia Arano

Citación recomendada: Silvia Arano. *Los tesauros y las ontologías en la Biblioteconomía y la Documentación* [en línea]. "Hipertext.net", núm. 3, 2005. <<http://www.hipertext.net>> [Consulta: 30 ene. 2007]. .

1. Introducción
2. Los tesauros
  - 2.1. Hacia una conceptualización de los tesauros
  - 2.2. Los tesauros y el entorno digital
3. Las ontologías
  - 3.1. Hacia una conceptualización de las ontologías
  - 3.2. Las ontologías y la Lingüística
4. Relación entre los conceptos de tesoro y ontología en el campo de la Biblioteconomía y la Documentación
5. Conclusiones
6. Bibliografía
7. Notas

## 1. Introducción

En los últimos tiempos, el ámbito de la representación y recuperación de información en la Documentación y la Biblioteconomía ha tenido que asumir el impacto de Internet y sus tecnologías asociadas, en especial el World Wide Web (WWW). Los cambios tecnológicos derivados de estos cambios tecnológicos están conduciendo a una progresiva digitalización del ámbito de la representación y recuperación de información que afecta por igual a los recursos de información, las herramientas de representación y recuperación, y los requerimientos de los usuarios.

A la luz de este contexto de creciente digitalización, son diversas las herramientas de representación y recuperación de la información que deben ser objeto de estudio, así como también son diversos los campos de conocimiento donde se estas herramientas se han originado: la Lingüística, la Inteligencia Artificial, la Documentación, la Ingeniería Lingüística... Así es que en la literatura especializada se analizan como herramientas para la representación y recuperación de información taxonomías, sistemas de clasificación, lexicones computacionales, bases de datos léxicas, tesauros, listas de encabezamiento, bases de conocimiento, mapas conceptuales, ontologías, anillos de sinónimos y redes semánticas, entre otros. Dentro de este amplio espectro de herramientas de representación y recuperación de información, son los tesauros y las ontologías los que con mayor frecuencia son objeto de vinculación en la bibliografía aunque provengan de ámbitos disciplinarios distintos. No obstante, la conceptualización que aplican los autores a los términos "tesoro" y "ontología" es muy diversa, y en ocasiones, los autores confunden, enfrentan, complementan o solapan ambos conceptos.

El objetivo general del presente artículo [1] es establecer la relación entre los conceptos de tesoro y de ontología para el campo de la Documentación y la Biblioteconomía. Con este fin se han fijado dos objetivos específicos. En primer lugar, analizar el concepto de tesoro a fin de delinear sus principales características y conocer sus semejanzas y diferencias con las ontologías. Y, en segundo lugar, precisar el concepto de ontología también a efectos de conocer sus características y analizar sus semejanzas y diferencias con los tesauros.

Sin embargo, no es nuestra intención normativizar tales conceptos sino exponer sus principales características, así como puntualizar sus semejanzas y diferencias para lograr una mejor comprensión de lo que es un tesoro y una ontología desde una perspectiva documental. O, dicho en palabras de Gilchrist (2002: 7):

"It would not be sensible to pontificate on the "correct" meanings of these words, but in trying to delineate the central characteristics, it is hoped that the reader will gain a clearer understanding of their differences and similarities"

## 2. Los tesauros

El término tesoro tiene su origen etimológico en la palabra latina *thesaurus*, que, a su vez, tiene su origen en la palabra griega *thesaurós*. En ambos casos, el significado era tesoro o repositorio de palabras.

En la época moderna, fue la Lexicografía el primer campo en el que se aplicó el término tesoro y en el que se elaboraron las primeras definiciones. En 1852 Peter Mark Roget publica su obra *ROGET'S THESAURUS of English Words and Phrases* (en lo sucesivo, *tesoro ROGET's*), cuya finalidad es la de proporcionar ayuda y asistencia a la expresión de ideas y a la composición literaria. El *tesoro ROGET's* tiene una organización conceptual, donde a partir de los conceptos explicitados en las entradas se relacionan y agrupan las palabras que designan o nombran, en diversos contextos, al concepto en cuestión. Es a partir de esta conceptualización propuesta por Roget que comienzan a diversificarse las acepciones y usos del término tesoro.

En 1957 se usa por primera vez el término tesauro en el campo de la Biblioteconomía y Documentación, en un trabajo presentado a la *Dorking Conference on Classification* por Helen Brownson, miembro de la *American National Science Foundation*. En dicha contribución, el término tesauro es utilizado al analizar la problemática de traducir los conceptos y sus relaciones, tal como se expresan en los documentos, a un lenguaje con mayor precisión y sin ambigüedades, para facilitar la recuperación de información. Plantear una evolución histórica de la concepción de los tesauros como herramienta documental en el campo de la la Biblioteconomía y Documentación, excede a los propósitos del presente artículo, y, por lo tanto, nos limitaremos a seleccionar algunas definiciones que, desde nuestro punto de vista, permiten delinear sus principales características.

## 2.1. Hacia una conceptualización de los tesauros

En el contexto normativo, *AENOR (Asociación Española de Normalización y Certificación)* define un tesauro en los siguientes términos:

" *Vocabulario de un lenguaje de indización controlado organizado formalmente con objeto de hacer explícitas las relaciones, a priori, entre conceptos (por ejemplo "más genérico que" o "más específico que")* ". (UNE 50-106-90, 1990: 5) [2]

Esta definición nos presenta al tesauro como una estructura organizada a partir de relaciones conceptuales explícitas, que incluye control de vocabulario, y cuya función es la indización.

Por otra parte, la norma elaborada por la *NISO (National Information Standards Organization)* [3], considera que un tesauro es:

" *a controlled vocabulary of terms in natural language that are designed for postcoordination* " ( ANSI/NISO Z39.19-2003, 2003: 1)

Esta definición agrega la característica del uso poscoordinado de los tesauros; o sea, la exigencia que sus términos sean interrelacionados en el momento de la búsqueda de información. Actualmente la *NISO* está culminando un proceso de actualización de dicha norma. Si bien todavía no contamos con el texto actualizado y definitivo, en la documentación preliminar que se encuentra disponible para consulta en el sitio web de la *NISO*, encontramos esta otra definición de tesauro que enfatiza la labor de control de vocabulario y la organización estructurada de dicha herramienta:

" *A set of word or phrases with equivalent terms explicitly identified and with ambiguous words or phrases (e.g.) homographs made unique. This set of terms also may include broader-narrower or other relationships* ". ( Z39.19 TAG Conference Call, 2003: 2)

Si consideramos las definiciones propuestas por los manuales de mayor difusión en el ámbito de la representación y recuperación de información, cabe destacar la de Aitchinson y Gilchrist (1987). Estos autores definen al tesauro como un vocabulario de un lenguaje controlado de indización, organizado formalmente de manera que las relaciones conceptuales son establecidas a priori, y que puede ser utilizado para la recuperación de información. Con esta definición, sumamos explícitamente otra de sus funciones: la recuperación de información.

Por su parte, Slype (1991) considera que el tesauro es:

" *una lista estructurada de conceptos, destinados a representar de manera unívoca el contenido de los documentos y de las consultas dentro de un sistema documental determinado, y a ayudar al usuario en la indización de los documentos y de las consultas* " (Slype, 1991: 24)

Este autor, si bien presenta características ya mencionadas (control de vocabulario, estructuración conceptual, uso en la indización y recuperación de información), amplía la definición al introducir al usuario como beneficiario de dicha herramienta.

A partir de las características que hemos resaltado de las definiciones anteriores, se puede establecer una primera conceptualización de lo que es un tesauro bajo nuestro punto de vista. [4] Un tesauro es "un tipo de lenguaje documental que representa la estructuración conceptual de un determinado campo del conocimiento. El tesauro proporciona una organización semántica principalmente a través de la explicitación de las relaciones establecidas entre dichos conceptos y, eventualmente, a través de un significado restringido de los términos que los representan. Los tesauros, para la Documentación y la Biblioteconomía, son por lo tanto una herramienta de control terminológico, puesto que la estructuración del tesauro es en base a conceptos. Los conceptos, son representados por términos seleccionados que evidencian ese control terminológico. Con dicho control se busca neutralizar la sinonimia y la polisemia, ambas características naturales de la lengua, que dificultan la precisión de la indización y recuperación de la información, ambas funciones básicas de los tesauros. Los tesauros son herramientas construidas para el auxilio tanto de los profesionales de la información como de los usuarios finales".

## 2.2. Los tesauros y el entorno digital

Como se ha mencionado en la introducción, el impacto tecnológico del entorno digital ha afectado por igual a recursos, herramientas y comportamiento del usuario en relación a la representación y recuperación de información. Esta influencia abre nuevas posibilidades en cuanto a las pautas de diseño, elaboración, gestión y uso de las herramientas utilizadas para la representación y recuperación de información. Desde esta perspectiva López-Huertas (1997), De la Cueva Martín (2000), Shiri y Revie (2000) y Qin y Paling (2001) entre otros, han empezado a perfilar los aportes del entorno digital a los tesauros, que podemos resumir en los siguientes elementos.

- Un primer elemento a tener en cuenta es el enriquecimiento de la funcionalidad de la estructura de los tesauros a partir de la hipertextualidad. Esto se traduce en el establecimiento de hipervínculos, entre todos los elementos estructurales (descriptores, no descriptores, notas de alcance, etc.), y también entre las distintas partes del tesoro.
- Un segundo elemento es la reducción de costos de actualización y mantenimiento. Debido a la creciente informatización de los procesos de construcción de tesauros y al progresivo abandono del soporte papel para la publicación de dichas herramientas, es viable una reducción de costos.
- Un tercer elemento, es la integración del usuario al proceso de creación, gestión y optimización de los tesauros, a través de tests de usabilidad, uso de técnicas de modelado de usuario, etc. Esto permite elaborar herramientas que tengan en cuenta los requerimientos de los usuarios, y descartar su construcción como simples estructuras teóricas.
- Un cuarto elemento es la posibilidad de aplicar medidas de reutilización e interoperabilidad en el momento de planificar y construir los tesauros. Con ello se posibilita el aprovechamiento y enriquecimiento de la información conceptual y lingüística ya generada para otros recursos.

Desde nuestro punto de vista, este último elemento es clave para la nueva generación de tesauros digitales. El uso de información conceptual y lingüística almacenada en otros tipos de recursos (por ejemplo, en una ontología) permite el enriquecimiento a nivel estructural de los tesauros, y posibilita el incremento de la amigabilidad de estas herramientas de cara a su utilización por parte de un usuario final no especializado.

## 3. Las ontologías

La Filosofía es el primer campo de conocimiento donde se utiliza el concepto de ontología. El uso de dicho concepto se origina en la noción aristotélica de *filosofía primera* -luego llamada *metafísica*-, que por un lado estudia la esencia del ser ( *el ser como ser* ), y por otro lado, estudia las características básicas de toda la realidad como un todo ( *el ser o el ente* principal del cual dependen los demás entes). En la actualidad, es considerada una rama de la Filosofía que tiene por objetivo la explicación de la existencia de una manera sistemática, trata sobre los tipos y estructuras de objetos, propiedades, eventos, procesos y relaciones relativas a cada porción de la realidad.

Es a partir de la década de los noventa, cuando las ontologías comienzan a tener protagonismo en la Inteligencia Artificial, con especial énfasis en la Ingeniería del conocimiento, el PLN y la Representación del conocimiento. En palabras de Studer (et al.), esto se debe a que:

*" Artificial Intelligence (AI) deals with reasoning about models of the world. Therefore, it is not strange that AI researchers adopted the term ontology to describe what can be (computationally) represented of the world in a program ".* (Studer et al, 1998: 25)

Por lo tanto, en un sentido muy general, las ontologías son para la Inteligencia Artificial recursos contruidos que permiten representar el conocimiento compartido y común sobre algo. Esta posibilidad de generar recursos compartibles y la consecuencia natural de intercambiar la información en ellos almacenada, es lo que provoca que un concepto como el de ontología (antes privativo de la Inteligencia Artificial) se filtre en los entornos de trabajo de otros ámbitos y en especial en lo relativo a la gestión de los recursos y herramientas del entorno digital. Si bien son diversos los campos de conocimiento [5] (Guarino 1998; McGuinness 2001) que refieren en sus investigaciones a las ontologías, es de nuestro interés analizar las ontologías como recursos lingüísticos y determinar la utilidad de su uso para la Documentación y la Biblioteconomía.

En el siguiente apartado se presentan una serie de definiciones del campo de la Inteligencia Artificial, de las cuales se extraen las principales características de la ontología como tal, con el fin de proponer una definición propia de ontología que guíe las consideraciones posteriores.

### 3.1. Hacia una conceptualización de las ontologías

A fin de aproximarnos hacia la comprensión del concepto de ontología, hemos optado por realizar una búsqueda y análisis de diversas definiciones de ontología en el ámbito de la Inteligencia Artificial. A continuación, se presentan las

definiciones seleccionadas y un breve comentario sobre las características que resaltan de las ontologías.

Para Neches et al. en 1991, una ontología se define de la siguiente forma:

" *an ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary* " (Neches et al. citado por Gómez-Pérez, 1999:33)

En esta definición, breve y concisa, se detallan sus partes componentes (términos, relaciones entre términos y reglas de combinación) e incluso se proponen elementos para su elaboración: identificar la terminología básica y las relaciones que se dan entre los términos; luego identificar las reglas que permitan combinarlos; y, por último, proveer las definiciones correspondientes a los términos y sus relaciones. La ontología se perfila como un recurso dinámico, en la medida que la constituyen términos creados a partir de reglas, además de los que son explícitamente definidos.

En cambio, Gruber propone la siguiente definición: "*an ontology is an explicit specification of a conceptualization* " (Gruber, 1993b: 1). El autor considera que una *conceptualización* está compuesta por objetos, conceptos y otras entidades que existen en una determinada área, y las relaciones que se dan entre ellos. Si bien la definición de Gruber es la más citada y difundida en la bibliografía sobre ontologías, es criticada por la noción de *conceptualización* [6] que utiliza, que es la misma que usan Genesereth y Nilsson en su manual "*Logical foundation of Artificial Intelligence*" de 1987. Para explicar el sentido de *especificación explícita*, se debe tener en cuenta que para la Inteligencia Artificial lo que existe es exactamente aquello que *puede ser representado computacionalmente*. Cuando el conocimiento de un dominio es representado en un formalismo declarativo, el conjunto de objetos que pueden ser representados es denominado *universo de discurso*. Este conjunto de objetos, y las relaciones entre ellos, se reflejan en el *vocabulario* con el que se representa conocimiento. En una ontología se asocian mediante definiciones los nombres de las entidades en el universo de discurso, (por ejemplo clases, relaciones, funciones u otros objetos) con un texto legible por el hombre que las describe y con axiomas que restringen la interpretación y otorgan la característica de *bien formados* a dichos términos. Formalmente, una ontología es la declaración de una teoría lógica.

En 1995, Guarino y Giaretta realizan un estudio en el que recopilan siete definiciones donde el concepto de ontología tiene distintas interpretaciones:

" 1. *Ontology as a philosophical discipline*; 2. *Ontology as an informal conceptual system*; 3. *Ontology as a formal semantic account*; 4. *Ontology as a specification of a conceptualization*; 5. *Ontology as a representation of a conceptual system via a logical theory*: 5.1 *characterized by specific formal properties*, 5.2 *characterized only by its specific purposes*; 6. *Ontology as the vocabulary used by a logical theory*; 7. *Ontology as a (meta-level) specification of a logical theory* " (Guarino y Giaretta, 1995: 1)

Estas definiciones (exceptuando la de contenido filosófico) pueden clasificarse en dos grupos: a) las que conciben la ontología como un marco conceptual a nivel semántico (definiciones 2 y 3); b) las que la conciben como un objeto concreto a nivel sintáctico con un uso guiado por un propósito específico (definiciones de la 4 a la 7). En definitiva, para Guarino una ontología es:

" *an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words* ". (Guarino, 1998: 2)

Se considera pues que estamos ante un recurso de la ingeniería compuesto por un vocabulario con un significado acotado (a través de presupuestos explícitos), cuya finalidad es la descripción de una porción de la realidad.

En 1996, Bernaras et al. proponen la siguiente definición:

" *An ontology provides the means for describing explicitly the conceptualization behind the knowledge represented in a knowledge base* ". (Bernaras et al. citado por Gómez-Pérez, 1999: 34)

Además de las características ya mencionadas en otras definiciones - *descripción explícita* de una *conceptualización*, que aporta el *significado* -, Bernaras et al. nuevamente hacen referencia a la ontología como recurso destinado a representar el conocimiento en una base de conocimiento.

Para Swartout et al. una ontología es:

" *a set of structured terms that describes some domain or topic. The idea is that an ontology provides a skeletal structure for a knowledge base* ". (Swartout et al. 1996).

Esta definición no aporta información nueva sobre las características que se han estado perfilando a partir de los demás autores citados, sin embargo es interesante rescatar el entorno del cual surge: la preocupación por compartir conocimiento entre sistemas a fin de reducir costos y dificultades y la conexión explícita como parte de una base de datos de conocimiento.

Uschold y Gruninger, en cambio entienden que:

" *Ontology is the term used to refer to the shared understanding of some domain of interest that may be used as a unifying framework to solve problems ... An ontology necessarily entails or embodies some sort of world view with respect to a given domain. The world view is often conceived as a set of concepts (e.g. entities, attributes, processes), their definitions and their inter-relationships; this is referred to as a conceptualisation* ". (Uschold y Gruninger, 1996: p. 5)

Estos autores también insisten sobre la naturaleza explícita de la representación que la ontología aporta sobre una conceptualización.

En 1997, Borst reformula la definición de Gruber precisando que "*Ontologies are defined as a formal specification of shared conceptualization*" (Borst citado por Gomez-Perez, 1999: p. 33). El aporte de dicha precisión es incluir la calificación de *compartida* a la noción de *conceptualización*, cuando en el carácter de *compartido* subyace una de las razones de la existencia de las ontologías como recursos. (Patil et al. 1992; Gruber 1993b; Guarino, Giaretta 1995).

Studer et al consideran que de las múltiples definiciones dadas sobre ontologías, la de Gruber (1993) y la de Borst (1997) son las que logran captar lo fundamental de dicho concepto, y las fusionan en la siguiente definición: "*An ontology is a formal, explicit specification of a shared conceptualisation*" (Studer et al, 1998: 25) Los elementos que componen esta definición, son explicados por los autores de la siguiente forma: a) *Conceptualización*: se refiere a un modelo abstracto de algún fenómeno en el mundo proveniente de haber identificado los conceptos relevantes de dicho fenómeno, b) *Explícita*: alude a que el tipo de concepto usado y las restricciones para su uso son explícitamente definidas, c) *Formal*: se refiere al hecho de que la ontología debería ser legible por ordenador, y d) *Compartida*: refleja la noción de que una ontología captura conocimiento consensual, que no es objeto de un solo individuo, sino aceptado por un grupo.

Las definiciones anteriormente expuestas, demuestran que existe una multiplicidad de interpretaciones posibles del concepto de ontología, que aportan distintos y complementarios puntos de vista, aún dentro de la misma área del conocimiento. Sin embargo, consideramos que ninguna se ajusta a todos los parámetros mencionados, y por lo tanto proponemos como definición sintetizadora la siguiente: una representación explícita y formal de una conceptualización compartida, que implica una perspectiva sobre cierta realidad y que se constituye en la estructura conceptual de una base de conocimiento. Asimismo su objetivo final, como se rescata de las definiciones estudiadas, es el compartir el conocimiento que representa.

### 3.2. Las ontologías y la Lingüística

En el apartado 3.1 propusimos una definición de ontología para el campo de la Inteligencia Artificial, que nuestro parecer es referencial para comprenderla desde otros ámbitos. Si bien la ontología desde esta perspectiva es aplicable al trabajo documental, consideramos que visualizar a la ontología como un recurso lingüístico tiene una vinculación aún más estrecha con las finalidades y las funciones de la Biblioteconomía y la Documentación.

Partimos, pues, de que una ontología proporciona una representación formal del conocimiento donde los conceptos, las relaciones y las restricciones conceptuales son explicitadas mediante formalismos en un determinado dominio. Si consideramos su aplicación en la Lingüística, la función más frecuente es de apoyo para los sistemas de Traducción Automática Basada en el Conocimiento y para la Terminología. En ambos casos, la ontología es una representación formal y explícita de la estructura conceptual del campo sobre el que se trabaja.

Por consiguiente, la ontología, para la Lingüística, es uno de los módulos asociados a una base de conocimiento donde su función es la de apoyo semántico para las palabras; es decir, las palabras son descritas como objetos lingüísticos en una base de datos léxica y son relacionadas con una jerarquía conceptual localizada en una ontología. En caso de que la ontología se aplique como soporte semántico para la Terminología, los términos son descritos en una base de datos terminológica y son posteriormente relacionados con la estructura conceptual de la ontología.

Entonces, la ontología desde un punto de vista lingüístico es un recurso aplicable al procesamiento documental puesto que facilita la estructuración conceptual de un campo del conocimiento determinado, que coincide en muchos aspectos con las representaciones de conocimiento utilizadas en la Biblioteconomía y la Documentación para la recuperación de información.

## 4. Relación entre los conceptos de tesauro y ontología en el campo de la Biblioteconomía y la Documentación

En los apartados 2 y 3 hemos delineado las principales características de los tesauros y ontologías, y hemos propuesto una conceptualización para cada una de estas herramientas de representación y recuperación de información. Del análisis realizado de la selección de definiciones, concluimos que:

- Un **tesauro** es una herramienta documental utilizada en el ámbito de la representación y recuperación de información, que representa un ámbito del conocimiento determinado mediante su estructuración conceptual. La estructura conceptual proporciona una organización semántica mediante la explicitación de las relaciones conceptuales y de la restricción del significado de los términos que los representan. El ámbito de conocimiento se estructura mediante relaciones conceptuales de tipo jerárquico, asociativo y de equivalencia. Un tesauro es utilizado tanto por usuarios profesionales de la información como por usuarios finales.
- Una **ontología** es una representación formal y explícita de la estructura conceptual de un campo del

conocimiento. Una ontología es el soporte semántico de las palabras que son descritas como objetos lingüísticos en una base de datos léxica o terminológica. Las relaciones conceptuales que son representadas en una ontología son muy variadas y dependientes del campo del conocimiento a estructurar. Una ontología se construye a fin de compartir y reutilizar la información almacenada, la cual, por estar formalizada puede ser interpretada tanto por personas como por programas informáticos.

Como también ya hemos indicado, si consideramos a la ontología como recurso lingüístico es clara su estrecha vinculación con la construcción de herramientas documentales para la representación y la recuperación de información.

La ontología aporta al proceso de construcción de herramientas documentales para la representación y la recuperación de información, la declaración explícita de las relaciones conceptuales en un ámbito en particular a través de la formalización semántica de dicha estructura.

La formalización semántica, a su vez, posibilita una representación lógica y coherente de dicha estructura conceptual, por lo tanto permite una interpretación informática que se orienta hacia la interoperabilidad y reutilización por parte de otros recursos o aplicaciones. La formalización semántica, por lo tanto, implica traducir a una sintaxis legible por ordenador el enunciado que aporte el significado del tipo de relación existente entre dos o más conceptos.

Como ejercicio de aplicación proponemos formalizar las relaciones jerárquicas de la faceta de un tesoro. Para este caso, tomaremos como ejemplo la faceta *CÉLULAS* del *Tesoro ICYT de Biología Animal del CINDOC*. De dicha faceta se analiza la relación jerárquica entre el concepto del encabezamiento de faceta y los conceptos del primer nivel de especificidad.

Figura 1 Faceta *CÉLULAS* [7] : extracción de encabezamiento de faceta y conceptos subordinados de primer nivel de especificidad

**CELULAS**  
> ADIPOCITOS  
> ASTROCITOS  
> CELULAS DE KUPFFER  
> CELULAS ENDOTELIALES  
> CELULAS EPITELIALES  
> CELULAS GONADALES  
> CELULAS MAMOTROPAS  
> CELULAS MESODERMICAS  
> CELULAS NEOPLASICAS  
> CELULAS SECRETORAS  
> CITOPLASMA  
> ENTEROCITOS  
> ERITROBLASTOS  
> HEPATOCITOS  
> LINEAS CELULARES  
> MACROFAGOS  
> MASTOCITOS  
> MEGACARIOCITOS  
> MEMBRANA PLASMATICA  
> MIOBLASTOS  
> NEURONAS  
> NUCLEO CELULAR  
> PARED CELULAR  
> PINEOCITOS  
> SINOVIOCITOS  
> ZIGOTO

En primer lugar, observemos la relación jerárquica entre el concepto "células" como término general con respecto a dos de los conceptos etiquetados como sus específicos (>), "citoplasma" y "adipocitos". La formalización de dicha relación permitiría notar que no son subordinados al concepto general a partir de la misma característica. Mientras que "citoplasma" se considera *una parte* de la célula (formalmente "citoplasma" <parte de> "células"), en realidad "adipocitos" es *un tipo* de célula. Por lo tanto, la formalización se correspondería "adipocitos" <es un> "células". Si bien estamos ante dos casos de relación jerárquica, no son del mismo tipo. La relación de "citoplasma" y "células" es una



relación de *todo/parte* , y la relación entre "adipocitos" y "células" es una relación de *género/especie* .

La confluencia de ambas características para la construcción de una misma faceta genera problemas de coherencia estructural, puesto que teóricamente una faceta debe construirse a partir de la aplicación de un único criterio de subdivisión a un concepto determinado. Pero este hecho no solamente afecta la construcción "purista" de tesauros, sino que también, en casos más complejos, puede guiar erróneamente en la búsqueda y/o recuperación de información a un usuario final.

Es por ello que la formalización semántica aportada por las ontologías tiene una clara utilización en la verificación de la coherencia de las relaciones conceptuales. Incluso vale la pena preguntarse si la ontologización de las relaciones conceptuales (en el contexto de los lenguajes documentales) posibilitaría también que el etiquetado de dichas relaciones pudiera trasladarse a la interfaz con el usuario haciendo más transparente la estructura conceptual al momento de la búsqueda y/o recuperación de información.

En síntesis, la semejanza entre ambas herramientas de representación y recuperación de información radica en que estructuran conceptualmente determinadas porciones de la realidad.

En cambio, la diferencia entre ambas herramientas de representación y recuperación de información se basa en el nivel de abstracción con el que se construye la organización conceptual. Las ontologías permiten un desarrollo semántico de mayor profundidad puesto que proporcionan una descripción lógica y formal de la información que almacenan, y por lo tanto puede ser interpretada tanto por usuarios humanos como por programas informáticos. Los tesauros, son herramientas con menor expresividad semántica y la información que estructuran solamente puede ser interpretada por usuarios humanos.

Para comprender un poco más esta diferencia de abstracción que planteamos, reflexionemos sobre lo que tendríamos en cuenta al momento de integrar dos conceptos, tanto a una estructura ontológica como a una estructura tesaural [8] .

En primer lugar, la integración en una estructura ontológica requiere que ambos conceptos sean definidos, que se establezca su pertenencia a las categorías conceptuales básicas propias de la ontología (p.e. objetos, eventos, propiedades, etc.), que se indique su posición en la jerarquía conceptual, que se indiquen sus relaciones conceptuales (que pueden variar según el área del conocimiento), y que se especifiquen sus propiedades y también toda la información que hereda de los conceptos jerárquicamente superiores a él. Cabe destacar que en una estructura ontológica, los atributos y las relaciones conceptuales pueden ser asignadas particularmente al concepto en cuestión, o también pueden ser heredadas de un concepto jerárquicamente superior. La ontología permite una herencia múltiple, es decir, que cada concepto puede recibir propiedades y relaciones conceptuales de más de un concepto superior en la jerarquía. También si a la ontología se le asocia una base de datos terminológica, tendríamos información sobre variantes y equivalentes lingüísticos para los términos relacionados a cada concepto. Además, todos estos parámetros de información especificados (definición, categorización, jerarquía, propiedades y herencia) son codificados de manera formal y explícita, o sea que no solamente serán entendibles mediante la lectura por parte de un humano, sino que pueden ser interpretados o descodificados por un programa informático y por lo tanto "manipulables" automáticamente.

En segundo lugar, la integración de ambos conceptos a una estructura tesaural requiere especificar principalmente, las relaciones conceptuales (restringidas a tres tipos: jerárquicas, asociativas o de equivalencia), en forma opcional una nota de alcance, y en forma excepcional una definición. A estos parámetros de información se les codifica mediante palabras, y tienen como lector destino a un humano. Asimismo, el establecimiento de las relaciones conceptuales es en cierta forma opaco, debido a que se indica el tipo de relación conceptual mediante un símbolo gráfico o abreviatura, y no se diferencian los subtipos dentro de cada tipo de relación conceptual (por ejemplo, en las relaciones jerárquicas, un subtipo serían las relaciones todo/parte y otro subtipo las de género/especie).

Estas diferencias de abstracción entre las estructuras ontológicas y las estructuras tesaurales, son a nuestra consideración las que paradójicamente muestran a la vez la cercanía y la distancia de los conceptos de ontología y tesoro.

## 5. Conclusiones

La creciente consolidación y expansión del entorno digital evidencia cada vez más la interrelación de los distintos ámbitos del conocimiento. En el caso de la Biblioteconomía y la Documentación, y más específicamente en el campo de la representación y recuperación de información es indudable la importancia de los aportes de la Lingüística y la Inteligencia Artificial.

Es en esta zona de intersección de la Biblioteconomía y la Documentación, la Lingüística y la Inteligencia Artificial donde es necesaria la retroalimentación entre los tesauros y las ontologías.

Nuestra reflexión debe partir de una reutilización de recursos, donde los esfuerzos en recopilar y sistematizar grandes volúmenes de información no se limite a la construcción de un recurso aislado, sino que dicha información sea estructurada de tal forma que pueda ser aprovechable para la generación o enriquecimiento de otros recursos. Es en este sentido, que proponemos el uso de una ontología, cuya estructuración permita formalizar un determinado dominio y ofrecer el soporte semántico necesario para proponer un modelo de tesoro. Esta vinculación posibilitaría un redimensionamiento de los tesauros como herramienta de representación y recuperación de información, con énfasis en

la coherencia y riqueza relacional de la estructuración conceptual.

En el presente artículo, solamente hemos ejemplificado una pequeña porción de la formalización semántica de las relaciones jerárquicas. Un campo de trabajo futuro sería la profundización en la formalización de las relaciones jerárquicas, como también la investigación de la formalización de las relaciones asociativas, de equivalencia o propias de un determinado ámbito del conocimiento.

Otro elemento a investigar sería la integración del concepto de "escenario", en tanto se analizarían las relaciones semánticas de un determinada estructura conceptual en función de los perfiles de usuario.

## 6. Bibliografía

ANSI/NISO Z39.19 - 2003 <i>Guidelines for the Construction, Format, and Management of Monolingual Thesauri</i> . [en línea] Bethesda: NISO. <a href="http://www.niso.org/standards/resources/Z39-19.pdf">http://www.niso.org/standards/resources/Z39-19.pdf</a> [Consulta: 19 marzo 2005]
A rano, S. (2004 ). "La ontología: una zona de interacción entre la Lingüística y la Documentación ". [en línea]. EN: <i>Hipertext.net</i> (2). <a href="http://www.hipertext.net">http://www.hipertext.net</a> [Consulta: 19 marzo 2005]
Barité, M. (1997). <i>Glosario sobre Organización y Representación del Conocimiento. Clasificación, Indización y Terminología</i> . Montevideo: Escuela Universitaria de Bibliotecología y Ciencias Afines.
Bechhofer, S.; Goble, C. (2001). "Thesaurus construction through knowledge representation". [en línea]. EN: <i>Data &amp; Knowledge Engineering</i> 37(2).25-45. <a href="http://www.sciencedirect.com/science">http://www.sciencedirect.com/science</a> [Consulta: 19 marzo 2005]
Cabré, M. T.; Feliu, J.; Vivaldi, J. (2004a). "Base de connaissances GENOMA: le rôle de l'ontologie". EN: <i>Journée d'étude Terminologie, Ontologie &amp; Representation des Connaissances</i> . Lyon: Université Jean-Moulin Lyon 3. 19-25.
Cabré, M. T.; Bach, C.; Estopà, R.; Feliu, J.; Martínez, G.; Vivaldi, J. (2004b). "The GENOME-KB project: towards the integration of concepts, terms, textual corpora and entities". EN: <i>LREC 2004, Fourth international conference on language resources and evaluation: held in memory of Antonio Zampolli</i> . Paris: ELRA. 87-90.
García, A. (2004). "Instrumentos de representación del conocimiento: tesauros versus ontologías". [en línea]. EN: <i>Anales de Documentación</i> (7). 79-95. <a href="http://www.um.es/fccd/anales/ad07/ad0706.pdf">http://www.um.es/fccd/anales/ad07/ad0706.pdf</a> [Consulta: 19 marzo 2005]
Gilchrist, A. (2002). "Thesauri, taxonomies and ontologies an etymological note". EN: <i>Journal of Documentation</i> 59(1). 7-18.
Gómez-Pérez, A. (1999). <i>Ontological engineering: a state of the art</i> . [en línea] Pennsylvania: School of Information Sciences and Technology(IST). Pennsylvania State University. <a href="http://citeseer.nj.nec.com/444416.html">http://citeseer.nj.nec.com/444416.html</a> [Consulta: 19 marzo 2005]
Gruber, T. R. (1993a). <i>A translation approach to portable ontology specifications</i> . [en línea] <a href="http://gicl.mcs.drexel.edu/people/regli/Classes/KBA/Readings/KSL-92-71.pdf">http://gicl.mcs.drexel.edu/people/regli/Classes/KBA/Readings/KSL-92-71.pdf</a> [Consulta: 19 marzo 2005]
Gruber, T. R. (1993b). <i>Toward principles for the design of ontologies used for knowledge sharing</i> . [en línea]. Pennsylvania: School of Information Sciences and Technology(IST). Pennsylvania State University. <a href="http://citeseer.nj.nec.com/gruber93toward.html">http://citeseer.nj.nec.com/gruber93toward.html</a> [Consulta: 19 marzo 2005]
Gruber, T. R. (2001). <i>What is an Ontology?</i> [en línea]. Stanford: Knowledge Systems Laboratory. Stanford University. <a href="http://www-ksl.stanford.edu/kst/what-is-an-ontology.html">http://www-ksl.stanford.edu/kst/what-is-an-ontology.html</a> [Consulta: 19 marzo 2005]
Guarino, N. (1998). <i>Formal ontology and information systems</i> . [en línea]. Trento, Roma: Laboratory for Applied Ontology (LOA). <a href="http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/FOIS98.pdf">http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/FOIS98.pdf</a> [Consulta: 19 marzo 2005]
Guarino, N. (1995). <i>Formal ontology, conceptual analysis and knowledge representation</i> . [en línea] Trento, Roma: Laboratory for Applied Ontology. <a href="http://www.loa-cnr.it/Papers/FormOntKR.pdf">http://www.loa-cnr.it/Papers/FormOntKR.pdf</a> [Consulta: 19 marzo 2005]
Guarino, N.; Giaretta, P. (1995). <i>Ontologies and knowledge bases: towards a terminological clarification</i> . [en línea]. Trento, Roma: Laboratory for Applied Ontology. <a href="http://www.loa-cnr.it/Papers/KBKS95.pdf">http://www.loa-cnr.it/Papers/KBKS95.pdf</a> [Consulta: 19 marzo 2005]



Guerrero, V.; Lozano, A. (1999). "Vínculos entre las Ontologías y la Biblioteconomía y Documentación". EN: <i>La Representación y la Organización del Conocimiento en sus distintas perspectivas: su influencia en la Recuperación de la Información (Actas del IV Congreso ISKO-España EOCONSID'99, 22-24 de abril de 1999 en Granada)</i> , 1999.Granada: ISKO-Facultad de Biblioteconomía y Documentación. 25-31.
ISO 2788:1986 (1986). <i>Documentation -- Guidelines for the establishment and development of monolingual thesauri</i> . 2 nd ed. Geneva: ISO.
López-Huertas, M. J. (1999). "Potencialidad evolutiva del tesoro: hacia una base de conocimiento experto". EN: <i>La Representación y la Organización del Conocimiento en sus distintas perspectivas: su influencia en la Recuperación de la Información (Actas del IV Congreso ISKO-España EOCONSID'99, 22-24 de abril de 1999 en Granada)</i> , 1999.Granada: ISKO-Facultad de Biblioteconomía y Documentación. 133-140.
López-Huertas, M. J. (1997). "Thesaurus structure design: a conceptual approach for improved interaction". EN: <i>Journal of Documentation</i> 53(2). 139-177.
McGuinness, D. L. (2002). <i>Ontologies come of age</i> . [en línea] Stanford: Stanford University. <a href="http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm">http://www.ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm</a> [Consulta: 19 marzo 2005]
Patil, R. et al. (1991). <i>The DARPA Knowledge sharing effort: progress report</i> . [en línea] Pennsylvania: School of Information Sciences and Technology(IST). Pennsylvania State University. <a href="http://citeseer.nj.nec.com/patil93darpa.html">http://citeseer.nj.nec.com/patil93darpa.html</a> [Consulta: 19 marzo 2005]
Qin, J.; Paling, S. (2001). "Converting a controlled vocabulary into an ontology: the case of GEM". [en línea]. EN: <i>Information Research</i> 6 (2). <a href="http://informationr.net/ir/6-2/paper94.html">http://informationr.net/ir/6-2/paper94.html</a> [Consulta: 19 marzo 2005]
Sánchez, E.; Martínez, R. (2002). <i>Las ontologías y su aplicación en el ámbito de la documentación</i> . [en línea]. València: Universitat Politècnica de València. <a href="http://ttt.upv.es/ccarrasc/doc/2000-2001/Ontolog%EDas/Index.htm">http://ttt.upv.es/ccarrasc/doc/2000-2001/Ontolog%EDas/Index.htm</a> [Consulta: 19 marzo 2005]
Shiri, A. A.; Revie, C. (2000). "Thesauri on the Web: current developments and trends". EN: <i>Online Information Review</i> 24(4). 273-279. <a href="http://dlist.sir.arizona.edu/archive/00000163/01/thesauri.pdf">http://dlist.sir.arizona.edu/archive/00000163/01/thesauri.pdf</a> [Consulta: 19 marzo 2005]
Slype, G. Van (1991). <i>Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales</i> . Madrid, Salamanca: Fundación Germán Sánchez Ruipérez.
Soergel, D. (1999). "The rise of ontologies or the reinvention of classification". EN: <i>Journal of the American Society for Information Science</i> (50). 1119-1120.
Studer, R. et al. (1998). <i>Knowledge engineering: principles and methods</i> . [en línea] Pennsylvania: School of Information Sciences and Technology(IST). Pennsylvania State University. <a href="http://citeseer.nj.nec.com/225099.html">http://citeseer.nj.nec.com/225099.html</a> [Consulta: 19 marzo 2005]
Swartout, B. et al. (1996). <i>Toward distributes use of large-scale ontologies</i> . [en línea] Calgary: University of Calgary. <a href="http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html">http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html</a> [Consulta: 19 marzo 2005]
Tudhope, D.; Alani, H.; Jones, Ch. (2001). "Augmenting thesaurus relationships: possibilities for retrieval". [en línea] EN: <i>Journal of Digital Information</i> 1(8). <a href="http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Tudhope/">http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Tudhope/</a> [Consulta: 19 marzo 2005]
Tramullas, J. (1999). "Agentes y ontologías para el tratamiento de la información: clasificación y recuperación en Internet". EN: <i>La Representación y la Organización del Conocimiento en sus distintas perspectivas: su influencia en la Recuperación de la Información (Actas del IV Congreso ISKO-España EOCONSID'99, 22-24 de abril de 1999 en Granada)</i> , 1999.Granada: ISKO-Facultad de Biblioteconomía y Documentación. 247-252
UNE 50106:1990. (1990). <i>Directrices para el establecimiento y desarrollo de tesauros monolingües</i> . Madrid: AENOR.
Uschold, M.; Gruninger, M. (1996). <i>Ontologies: principles, methods and applications</i> . [en línea] Pennsylvania: School of Information Sciences and Technology(IST). Pennsylvania State University. <a href="http://citeseer.nj.nec.com/uschold96ontologie.html">http://citeseer.nj.nec.com/uschold96ontologie.html</a> [Consulta: 19 marzo 2005]

Z39.19 TAG Conference Call, June 30, 2003. (2003). En: NISO. *Developing the next generation of standards for controlled vocabularies and thesauri*. [en línea] Bethesda: NISO. <http://www.niso.org/committees/MTinfo.html> [Consulta: 28 febrero 2005]

Silvia Arano forma parte del [Grupo DigiDoc](#) del [Instituto Universitario de Lingüística Aplicada](#) de la [Universidad Pompeu Fabra](#). Este artículo presenta una parte de los resultados del proyecto HUM2004-03162/FILO del Plan Nacional I+D+I del Ministerio de Educación y Ciencia (España)

## 7. Notas

[1] Artículo realizado en el marco del Proyecto financiado por el Ministerio de Educación y Ciencia, ref. HUM2004-03162/FILO "Web Semántica y Sistemas de información documental", coordinado por el Dr. Lluís Codina Bonilla (DIGIDOC - IULA UPF) [\[volver\]](#)

[2] Las normas UNE, editadas en España por la AENOR, son la traducción al castellano de las normas publicadas por la ISO ( International Organization for Standarization ) [\[volver\]](#)

[3] La NISO es una institución que desarrolla, mantiene y publica normas técnicas en el ámbito de la gestión de información a nivel de los Estados Unidos de América. [\[volver\]](#)

[4] Cabe aclarar que la citación de definiciones de tesoro no es exhaustiva, y dicha selección se encuentra guiada por las características que a nuestro criterio son relevantes en el concepto de tesoro como herramienta documental. [\[volver\]](#)

[5] Ingeniería del conocimiento, representación del conocimiento, modelado cualitativo, ingeniería del lenguaje, diseño de base de datos, modelado de información, integración de información, análisis orientado a objetos, extracción y recuperación de información, gestión del conocimiento, diseño de sistemas basados en agentes, integración empresarial, traducción del lenguaje natural, medicina, ingeniería mecánica, normalización de productos de conocimiento, recuperación y gestión de la información en WWW, bibliotecología y documentación, configuración y comercio electrónico. [\[volver\]](#)

[6] " according to Genesereth and Nilsson, a conceptualization is a set of extensional relations describing a particular state of affairs, while the notion we have in mind is an intensional one, namely something like a conceptual grid which we superimpose to various possible states of affairs." (Guarino; Giaretta 1995: 2) Según es interpretado por los autores citados, la limitación en la definición de Gruber estaría dada por no concebir como parte integrante de la conceptualización a la semántica de las relaciones, sino meramente a la existencia de dichas relaciones. Esto ocasiona que la definición de Gruber sea simplista, en el sentido que desconoce que una ontología explica una porción de una conceptualización, pero no podrá hacerlo de todas sus posibles dimensiones. [\[volver\]](#)

[7] Las marcas de estilo de letra cursiva y subrayado para los distintos conceptos subordinados son propias. Tienen por finalidad resaltar las diferencias conceptuales entre los conceptos subordinados al concepto que encabeza la faceta. [\[volver\]](#)

[8] En ambos casos se plantea una estructura conceptual de tipo genérico. [\[volver\]](#)